HIGH SCHOOL - DATA WAREHOUSE DESIGN

Business process

The Data warehouse is designed for the Students writing matura exam and receiving score business process. This process is described in the document Specification of business processes.

Relational Database schema

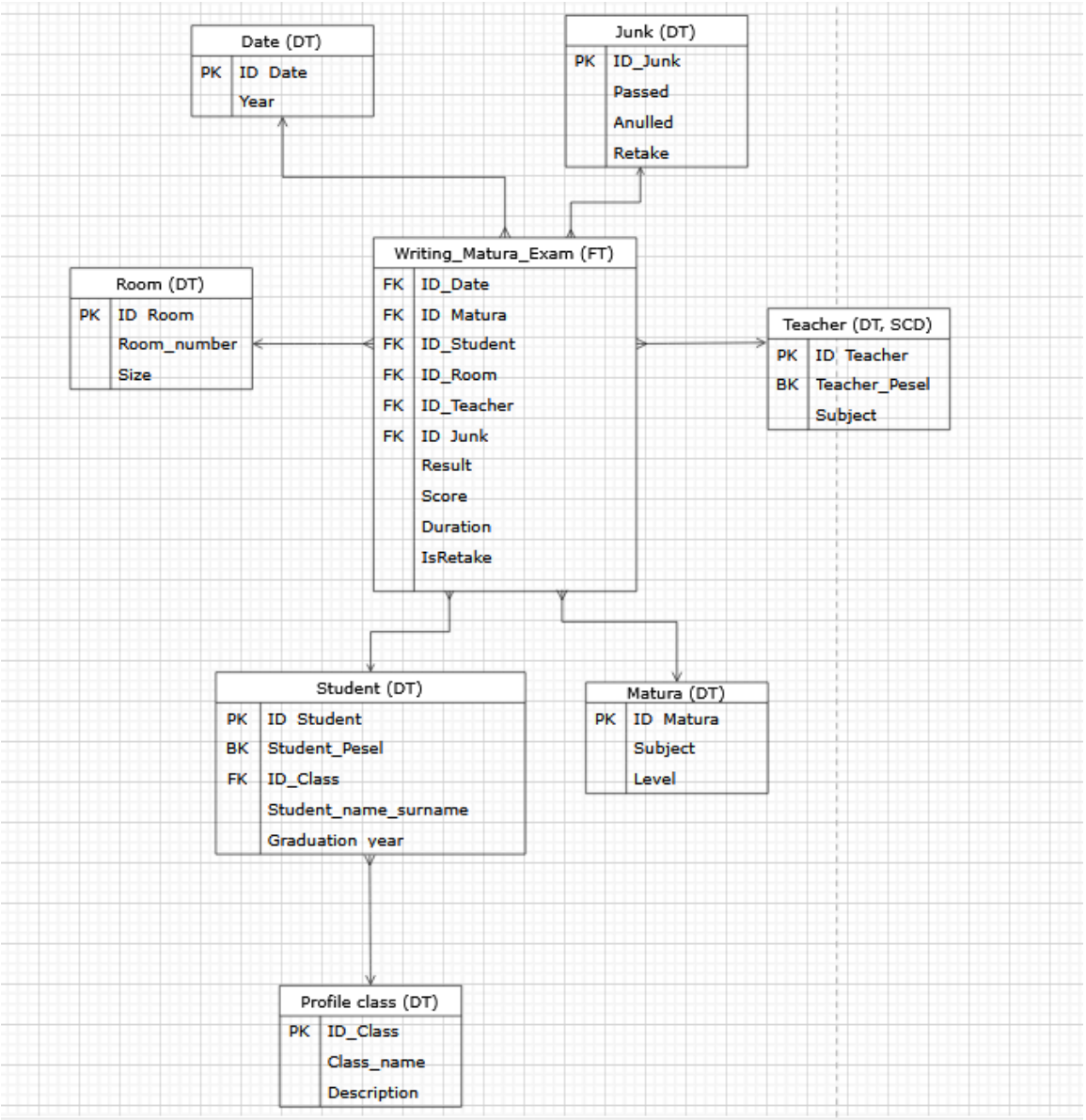

TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
WRITING_MATURA_EXAM (FACT TABLE)	one tuple describes one fact of writing the matura exam		
	ID_Date	Numeric	FK Date Date of the exam
	ID_Matura	Numeric	FK Matura Matura exam
	ID_Student	Numeric	FK Student Student writing exam
	ID_Room	Numeric	FK Room Classroom of the exam
	ID_Teacher	Numeric	FK Teacher Teacher supervising during the exam
	Result	Numeric	Result of the exam in decimal fraction. Allowed values: Number between 0 and 1, NULL
	Score	Numeric	Result * 100 Allowed values: between 0 and 100, NULL
	Duration	Numeric	Time (in min) a student takes to write an exam. Allowed values: number between 0 and 120, NULL
	IsRetake	Boolen	Information if the student was writing the retake Allowed values: 0 - Not retake, 1 - Retake
STUDENT (DIMENSION TABLE)	one tuple describes one student		
	ID_Student	Numeric	PK (surrogate key)
	Student_Pesel	11 digits	BK
	ID_Class	Numeric	FK Profile_Class Profile Class to which student belongs
	Student_name_surname	Varchar(40)	Student's name and surname
	Graduation_year	4 digits	Year of graduation
PROFILE CLASS (DIMENSION TABLE)	one tuple describes one profile class		
	ID_Class	Numeric	PK (surrogate key)
	Class_name	Varchar(15)	Name of the profile class. Allowed values: law, science, economics, medical, programming, biological, chemical, linguistic.
	Description	Varchar(50)	Information about extended subjects in that class. Allowed values: - Extended subjects: Biology, Chemistry; - Extended subjects: Polish, History; - Extended subjects: Mathematics, Physics; - Extended subjects: Mathematics, Geography;

			- Extended subjects: Mathematics, Informatics; - Extended subjects: Mathematics, Biology; - Extended subjects: Mathematics, Chemistry; - Extended subjects: Polish, English;
MATURA (DIMENSION TABLE)	one tuple describes one matura exam on a specified subject, level, and year of the exam		
	ID_Matura	Numeric	PK (surrogate key)
	Subject	Varchar(20)	Name of the subject. Allowed values: Mathematics, Polish, English, Physics, Informatics, Biology, Chemistry, Geography, History
	Level	Varchar(10)	Information about the level. Allowed values: basic, extended.
TEACHER (DIMENSION TABLE)	one tuple describes one teacher		
	ID_Teacher	Numeric	PK (surrogate key)
	Teacher_Pesel	11 digits	BK
	Subject	Varchar(20)	Name of the subject the teacher is teaching. Allowed values: Mathematics, Polish, English, Physics, Informatics, Biology, Chemistry, Geography, History
ROOM (DIMENSION TABLE)	one tuple describes one classroom at school		
	ID_Room	Numeric	PK (surrogate key)
	Room_Number	Varchar(4)	Number of the room
	Size	Varchar(5)	Information about the size of the room. Allowed values: small, big
DATE (DIMENSION TABLE)	one tuple represents the date of one matura exam		
	ID_Date	Numeric	PK (surrogate key)
	Year	4 digits	Year of the exam
JUNK (DIMENSION TABLE)	one tuple describes additional information about student's results		
	ID_Junk	Numeric	PK (surrogate key)
	Passed	Varchar(5)	Information if the student passed the exam. Allowed values: Passed, Not passed, NULL
	Annulled	Varchar(5)	Information if the exam was annulled. Allowed values: Annulled, Not annulled, NULL
	Retake	Varchar(10)	Information if the exam was a retake.

			Allowed values: Retake, Not retake

Dimensional model

Fact definitions

Fact 1 Writing matura exam: Matura exam from a specific subject on a specific level, written by a specific student from a specific profile class, on a specific day, in a specific classroom, supervised by a specific teacher. After two months of writing the matura exam, getting specific results from the exam.

Fact table: Writing_Matura_Exam

Granularity:

- Specified date of writing the matura exam
- Specified matura
- Specified student
- Specified room
- Specified teacher with specified subject that one is teaching
- Specified information if exam passed, if annulled and if it was a retake

Measures and aggregation functions:

Number of students - DISTINCT COUNT(ID_Student)

Sum of scores - SUM(Score)

Number of scores - COUNT(Score)

Average score calculated - sum of scores / number of scores

Number of classrooms - DISTINCT COUNT(ID_Room)

Number of retakes - SUM(IsReatke)

Number of present students - COUNT(Duration)

Dimension definitions

Dimensions for Fact 1 Writing matura exam fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
STUDENT	Student	Dimension
STUDENT PESEL	Student.Student_Pesel	Dimension attribute
STUDENT NAME AND SURNAME	Student.Student_name_surname	Dimension attribute
STUDENT YEAR OF GRADUATION	Student.Graduation_year	Dimension attribute
PROFILE CLASS	Profile_class	Dimension
PROFILE CLASS NAME	Profile_class.Class_name	Dimension attribute
PROFILE CLASS DESCRIPTION	Profile_class.Description	Dimension attribute
MATURA	Matura	Dimension
MATURA SUBJECT	Matura.Subject	Dimension attribute
MATURA LEVEL	Matura.Level	Dimension attribute
TEACHER	Teacher	Dimension
TEACHER PESEL	Teacher.Teacher_Pesel	Dimension attribute
TEACHER TEACHING SUBJECT	Teacher.Subject	Dimension attribute
TEACHER HIERARCHY	•Techer.Subject ••Teacher.Teacher_Pesel	Hierarchical dimension
ROOM	Room	Dimension
ROOM NUMBER	Room.Room_number	Dimension attribute
ROOM SIZE	Room.Size	Dimension attribute
DATE	Date	Dimension
EXAM YEAR	Date.Year	Dimension attribute
JUNK	Junk	Dimension
EXAM PASSED	Junk.Passed	Dimension attribute
EXAM ANNULLED	Junk.Annulled	Dimension attribute
EXAM RETAKE	Junk.Retake	Dimension attribute

Checking the feasibility of queries based on the multidimensional model

1. **Examine the trends in the number of students by the chosen level of matura exams across the last 4 years.**

   Measure: Number of students

   Dimension: Matura (dimension attribute: Matura level)

   Dimension: Date (dimension attribute: Exam year)

2. **Compare the results of students retaking the matura exam last year with the previous 3 years.**

   Measure: Average score

   Dimension: Date (dimension attribute: Exam year)

   Dimension: Retake (dimension attribute: Exam retake)

3. **Compare the average results of the students who wrote exams in smaller classrooms (up to 15 people) ang big classrooms (16-40 people) in the last year.**

   Measure: Average score

   Dimension: Room (dimension attribute: Room size)

   Dimension: Date (dimension attribute: Exam year)

4. **Compare the results from the matura exam from last year to the previous 3 years.**

   Measure: Average score

   Dimension: Date (dimension attribute: Exam year)

5. **Compare the results from extended math matura exams from students who were in a particular profiled class vs those who took the extended exam from the subject that was not extended in their class in the last 4 years.**

   Measure: Average score

   Dimension: Matura (dimension attribute: Matura level)

   Dimension: Matura (dimension attribute: Matura subject)

   Dimension: Profile Class (dimension attribute: Profile class description)

   Dimension Date (dimension attribute: Exam date)

6. **In how many small classes the matura exams were held last year compared to the previous 3 years?**

   Measure: Number of classrooms

   Dimension: Room (dimension attribute: Room size)

Dimension Date (dimension attribute: Exam year)

7. **How many people were writing other than extended math, extended maturas last year?**

   Measure: Number of students

   Dimension: Matura (dimension attribute: Matura level)

   Dimension: Matura (dimension attribute: Matura subject)

8. **How many people were writing extended math matura exams from certain profile classes last year?**

   Measure: Number of students

   Dimension: Matura (dimension attribute: Matura level)

   Dimension: Matura (dimension attribute: Matura subject)

   Dimension: Profile class (dimension attribute: Profile class name)

9. **How many people were retaking the extended math matura exam last year compared to two years ago?**

   Measure: Number of retakes

   Dimension: Matura (dimension attribute: Matura level)

   Dimension: Matura (dimension attribute: Matura subject)

   Dimension: Date (dimension attribute: Exam year)

10. **How many students were absent on an exam last year?**

    Measure: Number of students

    Measure: Number of present students

    Dimension: Matura (dimension attribute: Matura level)

    Dimension: Matura (dimension attribute: Matura subject)

11. **In how many classrooms and in which classrooms the matura was annulled 2 years ago?**

Measure: Number of classrooms

Dimension: Room (dimension attribute: Room number)

Dimension: Junk (dimension attribute: Exam annulled)

Checking if there are Data in the Data sources needed to fill the Data warehouse

TABLE NAME	COLUMN	SOURCE
WRITING_MATURA_EXAM (FACT TABLE)	one tuple describes one fact of writing the matura exam	
	ID_Date	Date ID. Foreign key from dimension table. Based on the Year in the Matura table, in the Relational Database source
	ID_Matura	Matura ID. Foreign key from dimension table. Based on the subject, level and year in the Matura table in the Relational Database source.
	ID_Student	Student ID. Foreign key from dimension table. Based on the student's pesel (BK) in the Student table in the Relational Database source.
	ID_Room	Room ID. Foreign key from dimension table. Based on the number of rooms in Sheet 1 source.
	ID_Teacher	Teacher ID. Foreign key from dimension table. Based on the teacher's pesel (BK) and subject in Sheet 1 source.
	ID_Junk	Junk ID. Foreign key from dimension table. Based on the Passed, Annulled and Retake in table Student-Matura-Result in the Relational Database source
	Result	Student's results from a specific exam. Taken from Result in the Student-Matura-Result in the Relational Database source.
	Score	Student's score from the exam, calculated Result * 100. Result taken from the Result column in the Writing_Matura_Exam table in the data warehouse.
	Duration	Time (in min) a student took to write the exam. Based on the Duration in the Student-Matura-Result in the Relational Database source.
	IsRetaken	Information if the student was writing the retake. Based on the IsRetaken in the Student-Matura-Result in the Relational Database source.
STUDENT (DIMENSION TABLE)	one tuple describes one student	
	ID_Student	Student ID. Surrogate key - generated by the database

	Student_Pesel	Student's Pesel. Taken from student_PESEL from Student table from Relational Database source.
	ID_Class	Profile Class ID. Foreign key from dimension table. Based on the data from Profile-Class table from Relational Database source.
	Student_name_surname	Student's name and surname. Taken from student_name_surname column in Student table in the Relational Database source.
	Graduation_year	Student's graduation year. Taken from the graduation_year column in the Student table in the Relational Database source.
PROFILE CLASS (DIMENSION TABLE)	one tuple describes one profile class	
	ID_Class	Profile Class ID. Surrogate key - generated by the database.
	Class_name	Profile Class name. Taken from class_name column from the Profile-Class table from Relational Database source.
	Description	Information about profile class extensions. Taken from column description from Profile-Class table from Relational Database source.
MATURA (DIMENSION TABLE)	one tuple describes one matura exam on a specified subject, level, and year of the exam	
	ID_Matura	Matura ID. Surrogate key - generated by the database.
	Subject	Name of the subject. Taken from the subject column from the Matura table from Relational Database source.
	Level	Information about matura level. Taken from the level column from the Matura table from Relational Database source.
TEACHER (DIMENSION TABLE)	one tuple describes one teacher with a specified subject. Implementation of SDC 2.	
	ID_Teacher	Teacher ID. Surrogate key - generated by the database.
	Teacher_Pesel	Teacher Pesel. Taken from column E from Sheet 1 source.
	Subject	Teacher's subject. Taken from column F from Sheet 1 source.
ROOM (DIMENSION TABLE)	one tuple describes one classroom at school	
	ID_Room	Room ID. Surrogate key - generated by the database.
	Room_Number	Number of the room. Taken from column C from Sheet 1 source.
	Size	Size of the room. Taken from column D from Sheet 1 source.

DATE (DIMENSION TABLE)	one tuple represents the date of one matura exam	
	ID_Date	Date ID. Surrogate key - generated by the database.
	Year	Matura year. Taken from the year column from Matura table from Relational Database source.
JUNK (DIMENSION TABLE)	one tuple describes additional information about student's results	
	ID_Junk	Junk ID. Surrogate key - generated by the database.
	Passed	Information if the student passed the exam. Taken from passed column from table Student-Matura-Result from Relational Database source.
	Annulled	Information if the exam was annulled. Taken from annulled column from table Student-Matura-Result from Relational Database source.
	Retake	Information if a student was writing a retake. Based on the IsRetaken column from Student-Matura-Result from Relational Database source. Allowed values: 0 - Not retake, 1 - Retake.